

A System to Assist Visually Disabled Musicians to Participate in Orchestras

Thomas Brophy, Flaithri Neff

Limerick Institute of Technology, Ireland

Christopher Noonan, Ian Pitt

University College Cork, Ireland

Donal Fitzpatrick

Dublin City University, Ireland

Thomas.Brophy@lit.ie, chrisnoonan9867511@gmail.com, dfitzpat@computing.dcu.ie,
ianp@cs.ucc.ie, Flaithri.Neff@lit.ie

Abstract

Music is an activity primarily associated with the capacity to hear, but in Western forms there is a significant emphasis on an orchestral musician's ability to see when rehearsing and performing. In both of these circumstances, the musician relies on being able to see the conductor's cues in order to translate the physical gestures into musical expression, feeling, tempo and musical emphases. The rehearsal environment is even more complicated in that there are often sporadic musical entry-points as the ensemble jumps between sections of a score. This paper outlines early-stage research which aims to alleviate some of the difficulties encountered by blind musicians in an orchestral environment.

Keywords

Blind, Orchestral Musician, Kinect, Auditory Feedback, Multimodal

Introduction

Music is an activity primarily associated with the capacity to hear, but in Western Art music there is a significant emphasis on an orchestral musician's ability to see when rehearsing and performing within conventional ensemble environments. In both of these circumstances, the musician relies on being able to see the conductor's cues in order to translate the physical gestures into musical expression, feeling, tempo and musical emphases. The rehearsal environment is even more complicated in that there are often sporadic musical entry-points as the ensemble jumps between sections of a score.

Four primary obstacles can be easily identified for the visually disabled musician in such circumstances:

1. Reading traditionally formatted music scores;
2. Staying apace with the often un-methodological jumping of bar numbers as required by the musical director during rehearsals;
3. Following the conductor's gestural cues during rehearsals and live performances;
4. Inputting edits and comments to the score as required.

The authors are developing a system that will relay relevant information to musicians using non-visual means, without impacting on an individual's freedom and ability to hear their own performance and that of their peers. The long-term goal of this research is to package an audio/haptic feedback system with conductor gesture recognition into a portable tablet-based design that will allow blind musicians to participate in traditional orchestral activities.

Preliminary studies on both an auditory feedback system and a gesture recognition system have been conducted thus far. This paper will describe these studies as well as discuss the future research avenues proposed for this project.

Discussion

Despite the wealth of literature in the domain of gesture-capture systems, there exists little comparative material which focuses on the orchestral conductor. Previous work has not been targeted at professional conductors for use in a rehearsal or live performance, but have rather tended to focus on allowing members of the public to control the tempo of prerecorded audio. The ‘Effektorium’ (Effektorium 2014) installation at the Mendelssohn-Bartholdy museum in Leipzig is an example of this. ‘Effektorium’ uses a modified baton to allow users to control the tempo of a piece of music. Similar to this is the ‘Virtual Conductor’ in Haus der Musik, Vienna that also allows users to alter the playback speed of prerecorded audio by moving an electronic baton.

Gesture Capture Prototype Development

Though the primary beneficiaries of any system produced as a result of this ongoing work are blind orchestral musicians, it has proven necessary to begin this effort by carrying out investigations into the nature of gestures employed by conductors and the best way to capture them. To this end, an initial prototype system has been devised which analyses the movements employed by the conductor to convey their wishes to the musicians under their direction. The following sections describe this work in more detail.

Initial Implementation

An event driven architecture was chosen for the gesture recognition portion of the project. This allows for any detection of movement, as opposed to a polling architecture that only detects gestures at specific intervals. The first implementation aimed to create a set of templates for all hand gestures defined during the initial research phase, excluding the tempo

gesture. Tracked joint data was passed to a recognition algorithm that compared this data with coordinates in the templates. When a match was found, an event was triggered.

GesturePak was used to capture the conductor's performance of different poses within each movement. Specific joints were specified for capture and tracking, as were the relevant axes for each gesture. A matched triggered an event in real-time that played a sound file specific to each gesture.

Improvements to the Gesture Capture System

On completion of the initial prototype further refinements were introduced. The number of poses for each gesture composition was reduced to facilitate the Kinect's frame-rate and resulting CPU load in order to improve accuracy. The Kinect's maximum frame-rate of 32fps proved impractical for all intended poses due to the quantity of data being processed when simultaneously tracking 20 joints on the skeletal frame.

Left hand gestures were added to the tracking process, with minimum and maximum duration times incorporated for storage in the XML gesture library. This allowed the system to acquire a more reliable interpretation of motions such as the 'quieter' gesture.

The tracking of tempo gestures proved very difficult to implement using the Kinect. This is primarily due to the variations between the expressions of each gesture, as well as the conductor's personalized performance. A single expression may have a considerable amount of variation due to subtle human inconsistencies, which impacts significantly on the tracking system's ability to adapt. Ultimately, machine-learning algorithms are probably required to improve accuracy in the long term, especially since this data relates specifically to each conductor's stylistic expressions. However, a short-term solution was implemented using a

tempo gesture boundary box, similar to the Mendelssohn Effektorium system. (Effektorium, 2014)

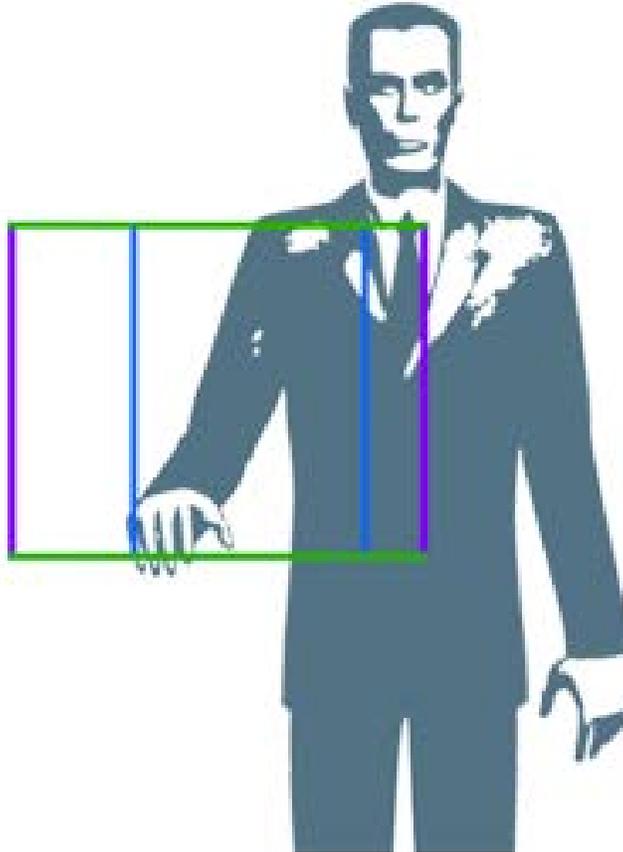


Fig.1. Boundary boxes for the Gesture Capture System. Green horizontal lines boundary pitch of gesture; light blue boundary width of gesture for soft expression; purple lines boundary loud expressions.

Referring to figure 1 above, the green horizontal lines indicate the pitch of a gesture. When the conductor's tempo gesture hand strikes one of these lines, the length of time between the current strike and last strike are calculated. Light blue vertical lines indicate the width of gesture using a soft expression/range of motion. When the width of the tempo gesture is between these lines, it indicates a small range of motion/expression, implying that the piece is to be played softly. Purple vertical lines indicate the width using a loud expression/range of

motion. When a tempo gesture has a range outside the light blue lines and inside the purple lines, it denotes that the range of motion is large and the indicated piece is to be played loudly.

Implementing the Tempo Gesture Aid

The first step in the implementation of the tempo gesture aid was to create a horizontal zone that output sound once the boundaries were reached. This was implemented using a method to draw bones and joints every time a movement is detected. An algorithmic comparison was created to compare the location of the right hand joint with the shoulder center joint on the y-axis. A minimum duration was incorporated to prevent constant retriggering while within the detection zone.

It became evident that the boundary box concept for tracking tempo gestures would be too far removed from a conductor's normal use of gestural movement. Therefore, three different zones were created in the x and y space. Zone 1 was used for minimal expression, zone 2 for medium expression, and zone 3 for maximum expression. Each zone, when entered, triggered a sound of different pitch and loudness to provide the most effective means of conveying aural information in a noisy environment (Walker et al, pp 2-7). Shoulder height was estimated to be an appropriate height to have the tempo gesture aid. Zone boundaries were scaled based on the distance between right shoulder joint and spine joint. Gesture movements were mapped within each zone boundary box. The zones are outlined in figure 2.



Fig.2. Zone boundaries for the Tempo Gesture Aid. Zone 1 is represented by an aqua colored line correlating with minimum expression; Zone 2 is represented by a purple colored line correlating with medium expression; Zone 3 is represented by a yellow colored line correlating with maximum expression.

Both auditory and visual feedback was relayed to users. Zone 1 displayed an aqua color line and played a low pitch sound for feedback. This range of motion indicated minimum expression (play softly). Zone 2 displayed a purple color line and played a tone that was higher in pitch than for zone 1, but lower than for zone 3. This range of motion indicated medium expression. Zone 3 triggered a pitch that was higher than the other two zones, and indicated the largest range of motion, implying that the conductor was expressing maximum volume.

Testing the Gesture Capture System

Testing was carried out with the assistance of two conductors. This test procedure comprised a Kinect sensor attached to a Windows laptop with an external display. The Kinect was positioned on top of the display. Gesture diagrams were placed in front of the display so that users could see them adequately.

Participants were familiarized with the system through a user-study information sheet and a short demonstration of the available functionality. A ‘User Guide’ document was produced in order to aid participants that described all aspects of how the system functioned, including detailed descriptions and diagrams of the gestures in the library. Descriptions of the user interface and of the different modes of system functionality were also explained. This user document was sent to the participants before the testing took place. This allowed the conductor to become familiar with the systems functionality and provided an understanding of the gestures in the library.

During the test procedure, participants were asked to experiment with the system while “thinking out loud” in Training Mode. They were asked to perform a series of gestures in order to control an embedded musical piece proficiently. This provided interesting preliminary data and helped the participant to feel more comfortable during the process of testing (Rubin & Chisnell, 2008). Once participants stated that they were familiar with the system, the user-study began. Participant interaction with the system was monitored, and after the testing phase, participants were asked to complete a short questionnaire.

Feedback & Recommendations

Feedback and recommendations from both conductors was critical in making further improvements to the system. Recommendations included:

- A consideration of the fact that during a performance the first beat is often missed by musicians. Although musicians anticipate this initial beat/click, it is quite difficult to synchronize immediately as there is no precursor or robust warning. This results in situations where musicians actually start on the second beat/click. This scenario was catered for in the system by implementing a ‘begin’ gesture to precede the use of the tempo gesture bar.
- Facial expression and ‘in-breath’ actions while conducting often display pivotal information cues for musicians. These are often subtle but easily interpreted by musicians. In addition, these expressions are also subtle signs that serve as a precursor to a gesture.
- Conductors felt that the height of the tempo gesture bar was too high and that it should be lowered from shoulder height to chest height for a more natural gesture position.

Auditory/Tactile Attention Cues

In addition to the development of the Gesture Capture System, the authors are developing a design framework to deliver auditory and tactile feedback to blind musicians, with the intention of replacing traditional visual information cues employed in the orchestral context.

The first stage of this research focuses on auditory and tactile attention mechanisms in noisy rehearsal environments. A series of studies have been conducted to investigate the efficiency of various sonic icons and haptic feedback designs.

The first of these studies (Brophy et al., 2015) investigated the effect of varying auditory parameters on participants’ reaction times. Two sound environments were used. In sound environment 1, participants were exposed to an acoustically chaotic environment comprising a recording of a youth orchestra rehearsal. This was relayed to participants using headphones

while also presenting to them different auditory cues designed with randomized intervals. Sound Environment 2 formed the benchmark environment comprising only auditory cues without the noisy background environment. After hearing an auditory cue, participants were required to press a key on a keyboard. There were five cues each varying in volume, timbre, or rhythm. All cues were 350ms and increasing linearly in volume for 345ms with 5ms linear decrease. All cues consisted of six harmonic components with a fundamental of 800Hz. Cue 1 changed only in amplitude. Cues 2 and 3 changed in timbre with cue 2 progressing from consonant to dissonant, and cue 3 progressing from dissonant to consonant. Cues 4 and 5 changed in rhythm, with cue 4 comprising even rhythm (70ms bursts) and cue 5 comprising randomized rhythm. Results showed no significant difference in reaction times across four of the stimuli, with cue 5 showing significantly slower reaction time.

The second study (Brophy et al., 2015) compared the reaction times of auditory, tactile, and audio-tactile combination stimuli. This was conducted using cue 1 from the previous study played over headphones and a vibration motor attached to the inner forearm for relaying the tactile stimulus. Three sound environments were used, increasing by 6 dB increments (Env.1, Env. 2 (+6dB), Env. 3 (+12dB)). Results from this study showed both vibration-only and audio-vibration combination cues achieved significantly quicker reaction times than the auditory-only stimulus across all sound environments. The reaction time for the auditory-only stimulus was also significantly faster in Env. 1 compared to either of the other sound environments.

Conclusions

At an early stage, technical issues became apparent whereby the Kinect was not capable of recognizing gestures beyond a certain level of complexity. However, the system was effective in its tracking and recognition of the set gestures based on the gesture library. Further

refinement to the library in order to personalize the system to a particular conductor's style was also fruitful when analysis during rehearsals was possible. Deviation from the gesture library without modification, however, remains problematic at this early stage of development. Furthermore, some gestures are inherently difficult to track accurately with the Kinect sensor, even when included in the library. For example, the 'louder' gesture involves raising the wrist from waist to shoulder height, but in order to repeat the gesture the wrist needs to be lowered to waist height, which is also an indication for the 'quieter' gesture. Issues also remain with regard to the accuracy of the skeletal tracking generally, and the refinement of this aspect is fundamental to retaining a reliable level of usability and limiting gesture recognition errors.

In relation to the auditory/tactile cueing system, results indicate that the use of tactile feedback seems to improve participant attention cueing across different background sound levels. However, more research is needed to discern the most efficient method of attracting attention in a noisy environment.

Works Cited

- Brophy, Tom et al. 'Evaluating The Cueing Efficiency Of Sound Icon Designs In A Multimodal System During Acoustically Chaotic Events'. The International Conference On The Multimodal Experience Of Music 2015. Sheffield, UK: 2015.
- Noonan, Chris. 'Making Conductor Gestures Accessible To Blind Musicians'. MSc. University College Cork, 2015.
- Rubin, Jeffrey, and Dana Chisnell. Handbook Of Usability Testing. Indianapolis, IN: Wiley Pub., 2008.
- Sokol, Zach. 'Control A Virtual Orchestra With Just The Flick Of Your Wrist | The Creators Project'. The Creators Project. N.p., 2015. Web: 20th July 2014.
- Walker, Bruce et al. 'High fidelity modeling and experimental evaluation of binaural bone conduction communication devices'. The 19th International Congress on Acoustics (ICA), Madrid, Spain: 2007.